



OPEN

Integrative competing endogenous RNA network analyses identify novel lncRNA and genes implicated in metastatic breast cancer

Dulari K. Jayarathna^{1,2}, Miguel E. Rentería^{2,3}, Jyotsna Batra^{1,3,4} & Neha S. Gandhi^{1,4}✉

Competing endogenous RNAs (ceRNAs) have gained attention in cancer research owing to their involvement in microRNA-mediated gene regulation. Previous studies have identified ceRNA networks of individual cancers. Nevertheless, none of these studies has investigated different cancer stages. We identify stage-specific ceRNAs in breast cancer using the cancer genome atlas data. Moreover, we investigate the molecular functions and prognostic ability of ceRNAs involved in stage I–IV networks. We identified differentially expressed candidate ceRNAs using edgeR and limma R packages. A three-step analysis was used to identify statistically significant ceRNAs of each stage. Survival analysis and functional enrichment analysis were conducted to identify molecular functions and prognostic ability. We found five genes and one long non-coding RNA unique to the stage IV ceRNA network. These genes have been described in previous breast cancer studies. Genes acted as ceRNAs are enriched in cancer-associated pathways. Two, three, and three microRNAs from stages I, II, and III were prognostic from the Kaplan–Meier survival analysis. Our results reveal a set of unique ceRNAs in metastatic breast cancer. Further experimental work is required to evaluate their role in metastasis. Moreover, identifying stage-specific ceRNAs will improve the understanding of personalised therapeutics in breast cancer.

Breast cancer (BC) is currently diagnosed in 1 in 8 Australian women over their lifetime, making it as the primary cause of female cancer-associated deaths in Australia. Decades of studies have identified candidate prognostic biomarkers for BC. Recent bioinformatic and experimental studies have found that microRNAs (miRNAs) can act as biomarkers in BC as they play a crucial role in transcriptional and post-transcriptional gene regulation. miRNAs belong to a group of small non-coding RNAs with 19–25 nucleotides in length. According to conventional RNA logic, miRNAs inhibit/degrade gene expression binding with miRNA response elements (MREs) of messenger RNAs (mRNAs)¹. Salmena, Poliseno² introduced the competing endogenous RNA (ceRNA) hypothesis revealing the bi-directional regulation mechanism of miRNAs. The ceRNA logic explains that non-coding RNA transcripts such as long non-coding RNAs (lncRNAs) with similar MREs can also bind with the relevant miRNAs modulating gene regulation and protein networks. Due to these reasons, ceRNAs have gained considerable attention in cancer studies.

Previous bioinformatic and experimental studies have identified many ceRNAs associated with BC risk³. Apart from generalising results for BC incidents, different cancer stages can also severely impact response to therapy and mortality. In BC, stage I refers to a tumour or small size confined to the breast; stage II explains the disease that has locally advanced beyond the breast; stage III describes BC has spread to the neighbouring organs, and stage IV refers to distant metastatic disease⁴. The early stages, I and II, are considered treatable compared to advanced stages, III and IV, that require more radical and active treatment strategies⁵. Therefore, identifying stage-specific biomarkers in BC will significantly contribute to understanding BC biology under different pathological states.

¹Centre for Genomics and Personalised Health, School of Chemistry and Physics, Queensland University of Technology, 2 George Street, Brisbane, QLD 4000, Australia. ²Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4006, Australia. ³School of Biomedical Sciences, Faculty of Health, Queensland University of Technology, Kelvin Grove, Brisbane, QLD 4059, Australia. ⁴Translational Research Institute, 37 Kent Street, Woolloongabba, QLD 4102, Australia. ✉email: neha.gandhi@qut.edu.au

In this study, we performed a ceRNA network analysis to identify stage-specific ceRNAs in BC, and this approach can be applied for any disease of interest in the future. The identified ceRNAs were further studied using two downstream analyses to identify their molecular functions and prognostic ability.

Materials and methods

Patients and samples collection. The expression data (RNA-seq and miRNA-seq) and clinical data of BC were collected from the cancer genome atlas (TCGA) that contains 1091 cases and 113 controls. The HTSeq-counts data of RNA-seq (including protein-coding and long non-coding) and isoform quantification data of miRNA-seq (in BC) were downloaded to a local computing server using the genomics data commons (GDC) data portal⁶. For selecting miRNA-seq data for individuals, a manifest file was generated using the GDC Data Portal. Then, the GDC data transfer tool was used to transfer data files listed in the manifest file. The differential expression analysis produced the "group" variable to identify differentially expressed protein-coding genes and long non-coding RNAs.

Differential expression analysis. Firstly, we removed TCGA BC samples with duplicated sample IDs. Then samples that are neither solid tissue normal nor primary tumour were removed as we compared primary tumour and healthy samples in the differential expression analysis. First, we performed the counts per million (CPM) normalisation to correct sample library size differences. The low-expressed genes that $\log(\text{CPM}) < 1$ in more than 50% of the samples were removed before the differential expression analysis⁷. Ignoring low-expressed genes improves the total count of differentially expressed genes and enhances sensitivity and precision. Raw counts expression data were re-normalised using the TMM (trimmed mean of M values) method implemented in the edgeR (3.40.0) R package (<https://bioconductor.org/packages/release/bioc/html/edgeR.html>) to compare expression levels between samples (excluding low-expressed genes)⁸. The normalised data were transformed into a standard scale using the voom function in the limma (3.54.0, <https://bioconductor.org/packages/release/bioc/html/limma.html>) (linear modelling for microarrays) R package⁹. Previous RNA-seq data analysis-related works have recommended this hybrid technique, TMM normalisation with voom transformation, due to its better performance in data preprocessing^{10,11}. Moreover, Oshlack et al. have shown TMM normalisation is robust and outperforms library size normalisation¹². In differential expression analysis, linear models were fitted for each gene using the "lmFit" function implemented in the limma (3.54.0, <https://bioconductor.org/packages/release/bioc/html/limma.html>) R package⁹. Then eBayes moderation was applied using information across all the genes to obtain more precise estimates of gene-wise variability. Four differential expression analyses were conducted for stage I to IV-control comparisons. The cancer stage was determined using the "pathologic stage" as it provides more accurate information combining results from clinical examinations and surgeries. We gathered 181 (19% basal-like, 5% HER2+, 62% luminal A, 13% luminal B, and 1% normal-like), 619 (21% basal-like, 13% HER2+, 40% luminal A, 25% luminal B, and 1% normal-like), 247 (12% basal-like, 16% HER2+, 40% luminal A, 29% luminal B, and 3% normal-like), and 20 (17% basal-like, 16% HER2+, 25% luminal A, and 42% luminal B) samples for stages I, II, III, and IV, respectively. In each stage-specific expression analysis, differentially expressed mRNAs, lncRNAs, and miRNAs were defined at $|\log_2\text{-fold change (FC)}| > 1$ and Benjamini-Hochberg (BH)-adjusted p value (default in limma package) < 0.05 ¹³.

Competing endogenous RNA network analysis. The differentially expressed mRNAs, lncRNAs, and miRNAs in each cancer stage were applied in the stage-specific ceRNA network analysis. The ceRNA network analysis consists of three main steps: (1) identifying lncRNA-mRNA pairs that share the significant number of miRNAs, (2) selecting positively correlated lncRNA-mRNA pairs, and (3) jointly estimating the significance of multiple miRNAs in lncRNA-mRNA pairs. These three steps are described in detail in previous ceRNA papers^{14,15}. The mRNA-miRNA and lncRNA-miRNA interactions are required to perform steps i and iii. We used miRcode and starBase databases for miRNA-target predictions^{16,17}. The miRcode database facilitates mRNA-miRNA and lncRNA-miRNA target predictions using a broad searchable map that contains 10,419 lncRNAs. The starBase includes miRNA-mRNA interactions predicted by analysing 108 CLIP-seq datasets. Steps i and ii were performed using the hypergeometric test and the Pearson correlation test, respectively. These two testing methods have been implemented in the GDCRNATools (1.18.0, <https://bioconductor.org/packages/release/bioc/html/GDCRNATools.html>) R/Bioconductor package¹⁸. The third step, multiple sensitivity correlation (mscor) analysis, was executed using the SPONGE (1.20.0, <https://bioconductor.org/packages/release/bioc/html/SPONGE.html>) (sparse partial correlation on gene expression) R/Bioconductor package¹⁹. The significant ceRNA interactions were filtered by three user-defined thresholds, (1) false discovery rate (FDR) < 0.01 in hypergeometric test, (2) Pearson correlation coefficient > 0.40 , and (3) the adjusted p value of mscor in SPONGE method < 0.05 . The resulting lncRNA-miRNA-mRNA associations in each BC stage were combined into a single column, as "<lncRNA gene ensemble ID>_<gene ensemble ID>_<miRNA name>". Then set of values in each BC stage were applied into a four-sets (for four stages) Venn diagram representation.

Functional enrichment analysis. A total of unique 47 aberrantly expressed genes (25 from stage I, 42 of stage II, 40 of stage III, and 47 of stage IV) were analysed to understand the biological functions of identified ceRNAs in this study. The Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) functional enrichment analyses were conducted using the R/Bioconductor clusterProfiler (4.6.0, <https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html>) R package²⁰.

Survival analysis. Survival analysis was performed using the Kaplan-Meier (K-M) survival curves, implemented in the survival (3.4.0, <https://cran.r-project.org/web/packages/survival/index.html>) R package to

explore the impact of the expression level of RNAs/miRNAs on the prognostic survival of patients²¹. For each gene/lncRNA/miRNA, the tumour samples were divided into two groups (low-expressed and high-expressed) according to the median expression level. The log-rank test (Mantel–Haenszel test) was used as the statistical method for the K–M curves. The log-rank test statistic has a chi-square (χ^2) distribution with one degree of freedom. Therefore, significant genes, lncRNAs and miRNAs were chosen under the χ^2 test statistic p value < 0.05. The survival-significant mRNAs were checked for tumour-suppressive/oncogenic/cancer-driven roles using the CancerMine database²².

Ethical approval. The study was approved by the Human Research Ethics Committees of the Queensland University of Technology (protocol code: 1900001147, date of approval: 19 December 2019) and the QIMR Berghofer Medical Research Institute (protocol code: P1051, date of approval: 23 August 2019).

Results

Differential expression analysis results. First, we conducted stagewise differential expression analysis to determine which genes, lncRNAs, and miRNAs are expressed at different levels between tumour and healthy groups. The number of up/downregulated lncRNAs, genes and miRNAs are available in Table 1.

The above-listed mRNAs, lncRNAs, and miRNAs were involved in the ceRNA network analysis.

Competing endogenous RNA networks of BC stages. We constructed four ceRNA networks for BC stages I–IV. In Table 2, we have included the count of significant ceRNA associations in each BC stage. The number of lncRNAs, genes, and miRNAs involved in stage-specific ceRNA networks is given within brackets.

According to Table 2, stage II and IV analyses have resulted in a considerably large set of ceRNA networks compared to stages I and III.

The detailed list of significant lncRNA-mRNA-miRNA associations of each BC stage is available in Supplementary Information. The Cytoscape tool version 3.9.1 (<https://cytoscape.org/download.html>)²³ was used to visualise ceRNA networks in each BC stage. Figures 1, 2, 3 and 4 illustrates ceRNA networks for BC stage I, II, III, and IV, respectively. In Figs. 1, 2, 3 and 4, blue-, green-, and yellow-coloured squares represent genes, lncRNAs, and miRNAs, respectively.

According to Figs. 1, 2, 3 and 4, hsa-miR-374a-5p and 374b-5p tend to build up separated ceRNA clusters in each BC stage. As shown in Fig. 4, the KLF5 gene and TRAF3IP2-AS1 lncRNA, unique to stage IV, create a unique triplet with hsa-miR-153-3p.

The lncRNAs, genes and miRNAs list of significant ceRNA associations were combined for a single variable as "<lncRNA gene ensemble ID>_<gene ensemble ID>_<miRNA name>" (Ex: ENSG00000234456_ENSG00000125845_hsa-miR-374b-5p). The values of gene ensemble ID, lncRNA ensemble ID, and the derived variable columns were inserted into a four-set Venn diagram representation (indicating four BC stages) as Fig. 5a–c, respectively.

According to Fig. 5, most lncRNAs, mRNAs, and lncRNA-mRNA-miRNA associations found in each BC stage ceRNA network have been shared among more than one stage. One lncRNA (TRAF3IP2-AS1) and five mRNAs (*KDR*, *SGCB*, *PRTG*, *KLF5*, and *PCNX1*) have been observed only in the stage-IV-specific ceRNA network. We identified 29 and 15 unique ceRNA associations in stage-III and stage-IV BC, respectively.

Breast cancer stage	lncRNA		mRNA		miRNA	
	Up	Down	Up	Down	Up	Down
I	50	45	1199	711	48	61
II	75	39	1385	779	62	67
III	71	43	1290	822	50	61
IV	90	43	1350	989	54	66

Table 1. Counts of differentially expressed (up/down) lncRNAs, mRNAs, and miRNAs in each BC stage. *lncRNA* long non-coding RNA, *mRNA* messenger RNA, *miRNA* microRNA, *Up* up-regulated, *Down* down-regulated.

Breast cancer stage	Number of ceRNA associations (number of involved unique lncRNAs, mRNAs, and miRNAs in the network)
I	48 (2, 25, 3)
II	127 (5, 42, 55)
III	86 (5, 40, 44)
IV	142 (6, 47, 59)

Table 2. The lncRNA-associated ceRNA networks for different breast cancer stages. *lncRNA* long-non-coding RNA, *ceRNA* competing endogenous RNA, *mRNA* messenger RNA, *miRNA* microRNA.

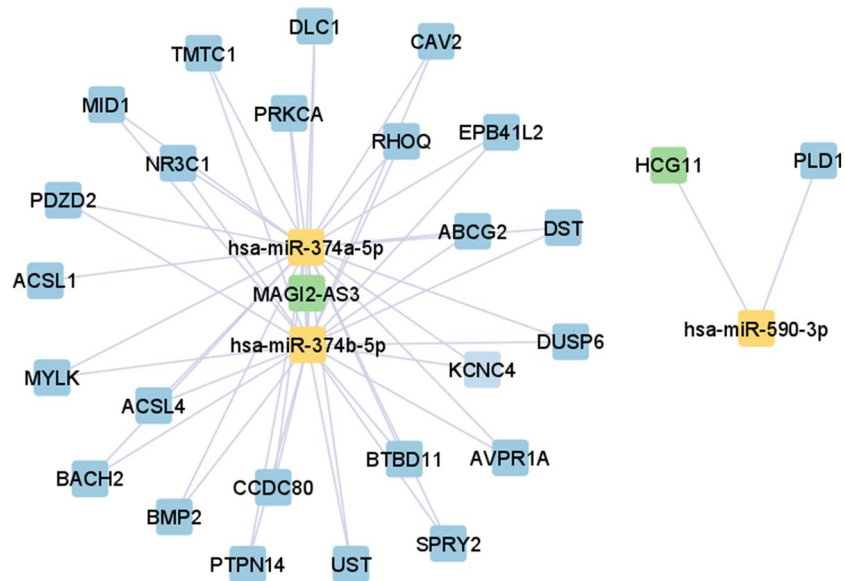


Figure 1. ceRNA network for stage I BC, was constructed by the Cytoscape tool²³.

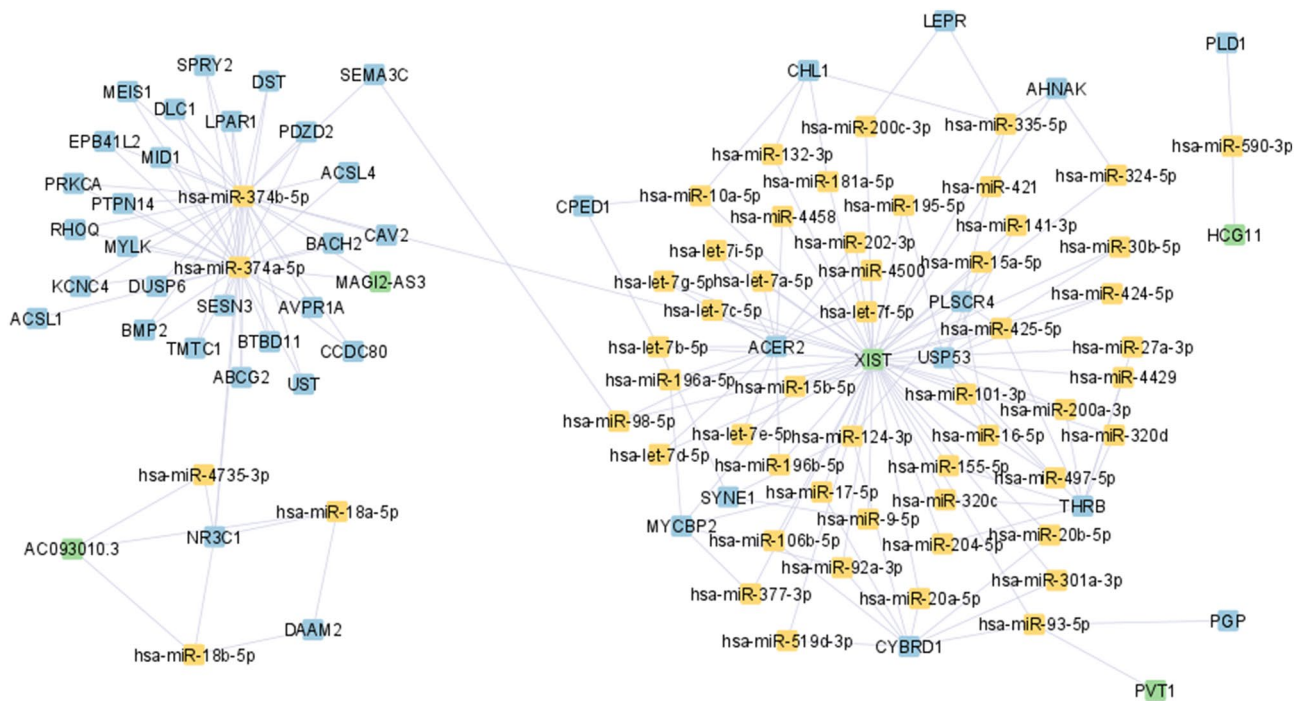


Figure 2. ceRNA network for stage II BC, was constructed by the Cytoscape tool²³.

Functional enrichment analysis for stage-specific competing endogenous RNA networks. We conducted functional enrichment analyses on 25, 42, 40, and 47 genes obtained from the stage I, stage II, stage III, and stage IV ceRNA networks. None of the stage-specific genes was enriched in KEGG pathways²⁴. Two genes resulting from stage I analysis were enriched in four CoA ligase activity-associated GO-molecular function (GO-MF) pathways. The GO-MF hormone-binding pathway was significant across stages II, III, and IV. In stage IV, six genes were enriched in three GO-cellular components (CC) pathways, membrane raft, membrane microdomain, and membrane region. Figure 6 illustrates GO pathway results for each BC stage.

Survival analysis for stage-specific competing endogenous RNA networks. K-M survival analyses and log-rank tests were performed to identify the potential stage-specific differentially expressed genes, lncRNAs and miRNAs strongly correlated with BC patients' prognostic characteristics. The significant genes, lncRNAs and miRNAs, were chosen under p value < 0.05. We found two, three, three, and one gene(s) as prog-

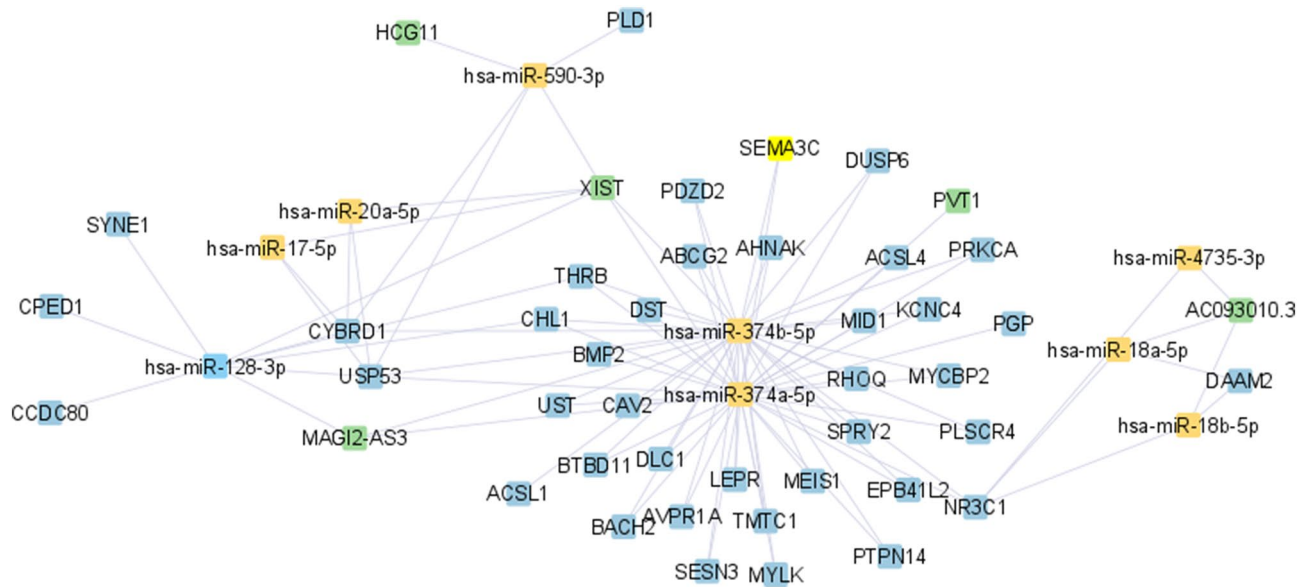


Figure 3. ceRNA network for stage III BC, was constructed by the Cytoscape tool²³.

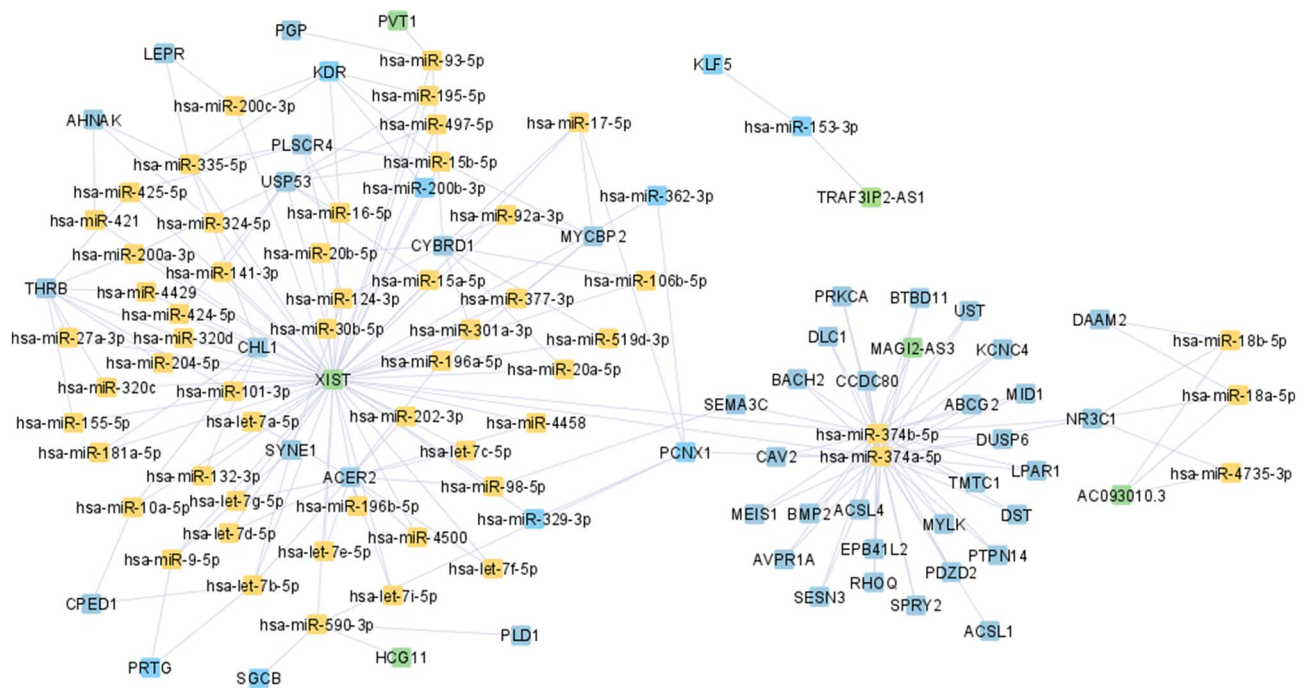


Figure 4. ceRNA network for stage IV BC, was constructed by the Cytoscape tool²³.

nostic biomarkers in stage I, stage II, stage III, and stage IV, respectively and described in Table 3. None of the lncRNAs was identified as biomarkers in stage-specific BC survival analysis. In miRNA-based survival analyses, two, three, and three miRNAs were statistically significant in stage I, stage II, and III, respectively. Figure 7 illustrates K–M curves for the top significant miRNA of each stage, hsa-miR-106b-5p in stage I, hsa-miR-31-5p in stage II, and hsa-miR-551b-3p in stage III.

Discussion

This study identified stage-specific lncRNA–mRNA–miRNA ceRNA associations in BC. According to ceRNA network analysis results, most ceRNA associations were shared across all four stages. In contrast, one lncRNA (TRAF3IP2-AS1) and five genes (*KDR*, *PRTG*, *KLF5*, *SGCB*, and *PCNX1*) were statistically significant only in metastatic BC, i.e., stage IV. The lncRNA TRAF3IP2-AS1 has been previously reported in renal cell carcinoma and glioblastoma but not in BC²⁵.

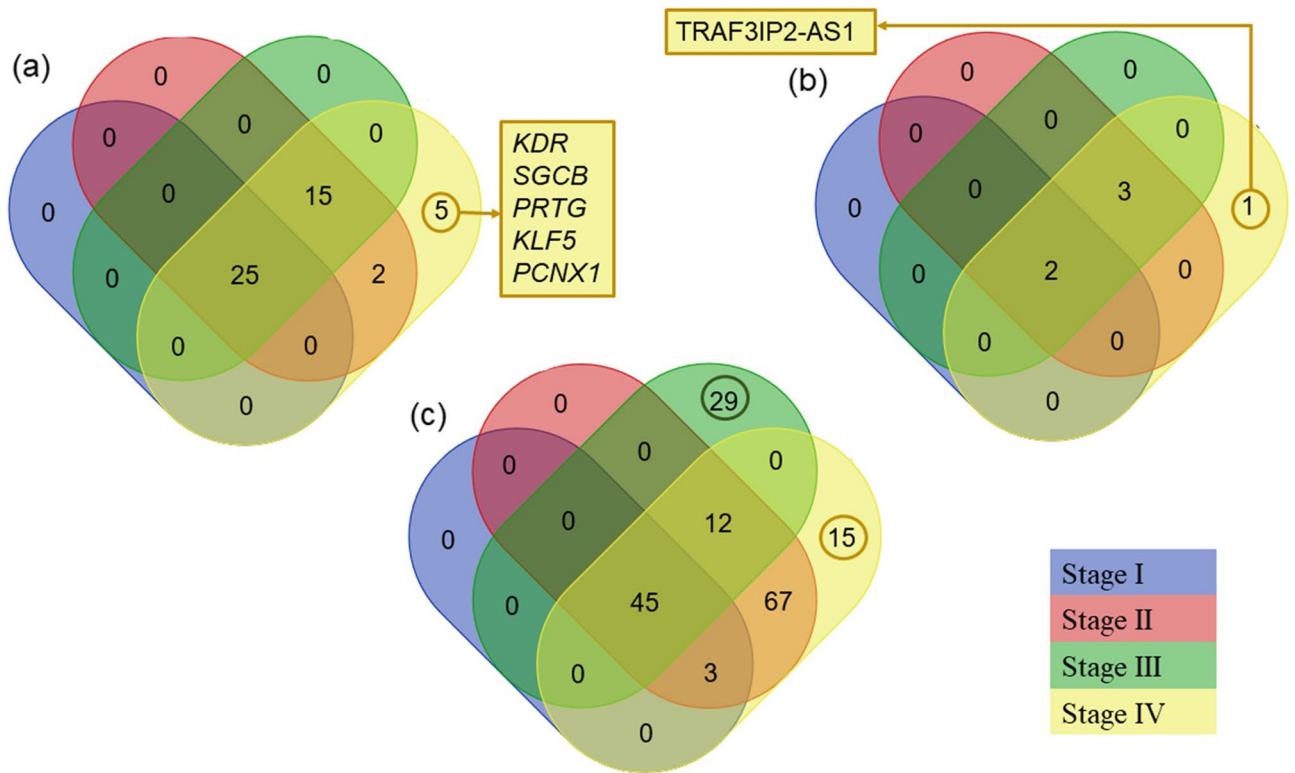


Figure 5. Venn diagram representation for mRNAs (a), lncRNAs (b), and competing endogenous RNAs (ceRNAs) (c) included in stage-specific significant ceRNA associations. Blue, Red, green, and yellow-coloured sets represent stage I, stage II, stage III, and stage IV, respectively. Most of the mRNAs and lncRNAs have been shared among four stages. Five genes, *KDR*, *SGCB*, *PRTG*, *KLF5*, and *PCNX1* and one lncRNA, *TRAF3IP2-AS1*, were observed only in stage IV, metastatic stage.

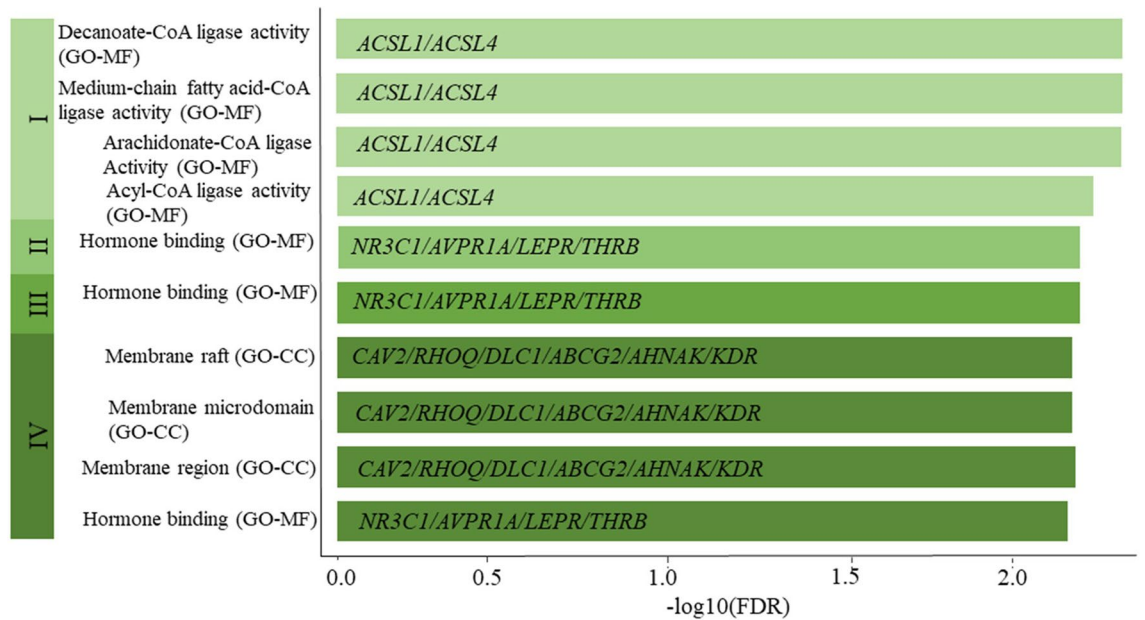


Figure 6. Pathway enrichment analysis results for genes included in stage-specific competing endogenous RNA networks in breast cancer²⁰. Four, one, one, and four gene-ontology pathways were statistically significant in stages I, II, III, and IV, respectively. The CoA ligase activity-related pathways have been significant only in stage I. The hormone-binding pathway was significant among stages II, III, and IV.

Stage	Gene	HR	p-value
I	<i>BMP2</i>	0.2934	0.0233
I	<i>BACH2</i>	0.2817	0.0185
II	<i>DUSP6</i>	0.5397	0.0122
II	<i>ACSL1</i>	1.8924	0.0111
II	<i>MYCBP2</i>	1.6608	0.0388
III	<i>DST</i>	0.5074	0.0234
III	<i>PDZD2</i>	1.9969	0.0291
III	<i>AHNAK</i>	0.2718	0.0221
IV	<i>MID1</i>	0.0895	0.0143

Table 3. Statistically significant genes from stage-specific survival analyses. *HR* hazard ratio.

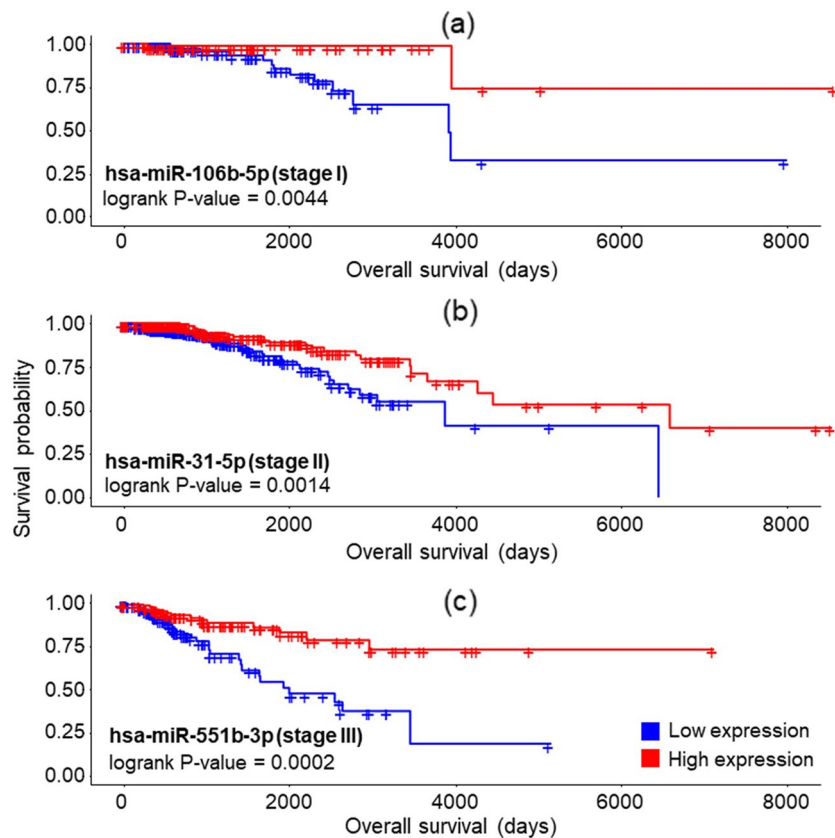


Figure 7. Kaplan–Meier survival plots for the top significant miRNAs in stage-specific breast cancer networks²¹. The high-expressed hsa-miR-106b-5p (a), hsa-miR-31-5p (b), and hsa-miR-551b-3p (c) were prognostic in stage I, II, and III, respectively. None of the microRNAs in the stage IV ceRNA network was significant from survival analysis.

According to AACR (American Association for Cancer Research) project GENIE (Genomics Evidence Neoplasia Information Exchange), the *KDR* gene is altered in 1.54% of BC patients. It can play an essential role in mediating endothelial cells proliferation, migration, and permeability²⁶. Endothelial cells are actively involved in cancer metastasis²⁷. The *PRTG* gene has not been previously described in BC. Nevertheless, *PRTG* has been identified as an oncogenic protein in gastric carcinogenesis by activating the downstream cGMP/PKG signaling pathway²⁸.

The *KLF5* gene is found to play a role in BC, but its precise function remains determined. On the one hand, the *KLF5* locus at chromosome 13 is frequently deleted in human BC, and its protein is degraded by the WWP1 oncogenic ubiquitin E3 ligase, which suggests a tumour-suppressor function²⁹. On the other hand, increased expression of *KLF5* is associated with expression of the HER2 oncoprotein and shorter survival in BC patients suggesting an oncogenic function of *KLF5* in BC³⁰. A recent bioinformatic study has found that SGCB protein is specific in basal A subtyped BC gene regulatory networks³¹. The *PCNX1* has been a potential marker of response

to chemotherapy in BC, and therapeutic modulation of its activities could enhance chemotherapy responses³². The BC studies mentioned above have described four out of five genes found from our stage IV-specific ceRNA network, and none of these studies explains their contribution to metastatic BC. Therefore, wet-lab experiments will be carried out in the future to investigate their role in BC metastatic nature.

Two downstream analyses, functional enrichment analysis and survival analysis, have ensured stage-specific ceRNA components found in our study. All statistically significant genes resulting from stagewise survival analyses have been previously reported in cancer studies²². *BMP2* from the stage I survival analysis has shown oncogenic function in BC³³. Only *DUSP6* and *ACSL1* in stage II have acted as an oncogene and a tumour suppressor in BC, respectively^{34,35}. *DST* and *AHNAK* genes that were significant from the stage III survival analysis have shown tumour suppressive characteristics in both experimental and computational BC studies^{36–40}. In stage IV, we identified *MIDI* as the only survival significant gene and it has been linked with invasive lobular carcinoma²². The invasive lobular carcinoma is the second most common type of BC. It originates in the milk-producing gland (lobules) of the breast. Invasive cancer is recognised as the cancer cells have broken out of the lobules where they initiated and are potential to expand to the lymph nodes and other areas of the body, leading to metastasis⁴¹. Therefore, *MIDI* gene should be further investigated to identify its role in metastatic BC (stage IV).

We found eight miRNAs from the stage-specific survival analyses, and these miRNAs have shown a tumour-suppressive/oncogenic role in previous BC experimental studies. The hsa-miR-106b-5p and hsa-miR-374b-5p were associated with the prognosis of stage I BC patients. The hsa-miR-106b-5p promoted cell migration, invasion, and proliferation by targeting *FUT6*⁴². Abnormal hsa-miR-374b-5p expression in luminal-HER2-positive BC cells can be used for classifying clinicopathologic subtypes of BC⁴³. Three miRNAs, hsa-miR-150-5p and hsa-miR-31-5p, are significant in stage II BC survival analysis, and hsa-miR-374b-5p was significant in both stages I and II. The other two miRNAs, hsa-miR-150-5p and hsa-miR-31-5p, have shown an oncogenic and tumour-suppressive role in BC, respectively^{44,45}. We identified three miRNA signatures (hsa-miR-551b-3p, hsa-miR-101-3p, and hsa-miR-26a-5p) as prognostics in stage III BC survival. Among them, hsa-miR-551b-3p have promoted oncogenic features in BC cells⁴⁶. In previous BC experimental studies, both hsa-miR-101-3p and hsa-miR-26a-5p have shown a tumour-suppressive role^{47,48}. We did not find statistically significant prognostic miRNAs from the stage IV BC survival analysis.

Functional enrichment analyses found molecular pathways associated with protein-coding genes in stage-specific ceRNA networks. Two genes in Acyl-CoA Synthetase Long (ACSL) Chain family, *ACSL1* and *ACSL4*, included only in stage I ceRNA networks, are enriched in four CoA ligase activity-associated pathways. Therefore, wet-lab experiments are required to understand the tumour-suppressive/oncogenic/cancer-driven role of ACSL Chain family members among early-stage BC patients. Four genes found in stages II, III, and IV (*NR3C1*, *AVPR1A*, *LEPR*, and *THRB*) are associated with hormone binding, which plays a role in BC pathophysiology and defining risk. Six genes in the stage IV ceRNA network are enriched in three components in the GO-CC pathway: membrane raft, membrane microdomain, and membrane region. These membrane domains have shown an important role in cancer metastasis⁴⁹.

Our study has shared a limited set of ceRNAs with the previous ceRNA network study for overall BC cases by Tuersong et al.³. Four miRNAs (hsa-miR-141, hsa-miR-200a, hsa-miR-204, and hsa-miR-301b) have been identified in ceRNA networks in both studies. Among these four miRNAs, Tuersong et al.³ have demonstrated that hsa-miR-204 was downregulated and hsa-miR-301b was upregulated in patients with BRCA compared with healthy controls and were associated with overall survival. Our previous transcriptome-wide association study also demonstrated hsa-miR-204 as a tumour-suppressive miRNA in prostate cancer with statistically significant low expressed levels in prostate cancer cell lines⁵⁰. Moreover, we found two genes, *SPRY2* and *CHL1*, involved in Tuersong et al.³ and our works (*SPRY2* in stages I–IV and *CHL1* in stages II, III, and IV). Observing a smaller number of shared ceRNAs between studies can be occurred due to higher heterogeneity between breast cancer stages, and it will lead to different RNA/gene expression levels. Zhou et al. conducted a ceRNA network analysis on BRCA subtypes, basal-like, HER2+, luminal A, and luminal B⁵¹. The authors have identified three lncRNAs, *NEAT1*, *OPI5-AS1*, and *AC008124.1*, among all four subtype-related ceRNA networks. Moreover, three lncRNAs, *NEAT1*, *FAM83H-AS1*, and *XIST1*, were significantly differentially expressed in the basal-like subtype-related network. Nevertheless, we could not find a shared outcome between our study and subtype-related networks. This can be due to our stage-based analyses containing RNA/gene expression levels from multiple subtypes.

This study is limited to ceRNA networks mediated by microRNA expression levels. Other genomic (copy number alteration), transcriptomic (transcription factors), and epigenetic (DNA methylation) factors were not considered in the ceRNA network analysis⁵². Moreover, other possible ceRNA components such as pseudogenes and lincRNAs were not considered. Therefore, future studies should be extended to address these concerns. Nevertheless, this study elucidates a new level of ceRNA network analysis, stage-specific ceRNA networks, to understand better common/unique ceRNA(s) among/within the stage(s) of a given cancer. Identifying novel stage-level cancer biomarkers will significantly contribute to the knowledge of personalised therapeutics and determining risk.

Conclusions

We conducted ceRNA networks analyses in four stages of BC. Only one lncRNA and five genes were significant in the stage IV BC ceRNA network. Further validation experiments are required to characterise their role in BC metastatic nature. Identifying ceRNA components across cancer stages will advance the diagnosis, risk identification, and therapeutics.

Data availability

Publicly available TCGA BC RNA-seq and miRNA-seq expression data were downloaded through the GDC Data Portal (<https://portal.gdc.cancer.gov/repository>). All statistical analyses and graph preparations were performed using the R statistical software, freely available at <https://cran.r-project.org/>. The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Received: 4 October 2022; Accepted: 7 February 2023

Published online: 10 February 2023

References

- Bhaskaran, M. & Mohan, M. MicroRNAs: History, biogenesis, and their evolving role in animal development and disease. *Vet. Pathol.* **51**(4), 759–774 (2014).
- Salmena, L. *et al.* A ceRNA hypothesis: The Rosetta Stone of a hidden RNA language?. *Cell* **146**(3), 353–358 (2011).
- Tuersong, T. *et al.* Comprehensive analysis of the aberrantly expressed lncRNA-associated ceRNA network in breast cancer. *Mol. Med. Rep.* **19**(6), 4697–4710 (2019).
- Sharma, G. N. *et al.* Various types and management of breast cancer: An overview. *J. Adv. Pharm. Technol. Res.* **1**(2), 109–126 (2010).
- Kwapisz, D. Oligometastatic breast cancer. *Breast Cancer* **26**(2), 138–146 (2019).
- Grossman, R. L. *et al.* Toward a shared vision for cancer genomic data. *N. Engl. J. Med.* **375**(12), 1109–1112 (2016).
- Liu, J. *et al.* Regulation of long non-coding RNA KCNQ1OT1 network in colorectal cancer immunity. *Front. Genet.* **12**, 684002 (2021).
- Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**(1), 139–140 (2010).
- Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**(7), e47 (2015).
- Morovat, *et al.* Identification of potentially functional circular RNAs hsa_circ_0070934 and hsa_circ_0004315 as prognostic factors of hepatocellular carcinoma by integrated bioinformatics analysis. *Sci. Rep.* **12**, 4933 (2022).
- Berti, F. C. B. *et al.* Comprehensive analysis of ceRNA networks in HPV16- and HPV18-mediated cervical cancers reveals XIST as a pivotal competing endogenous RNA. *Biochim. Biophys. Acta Mol. Basis Dis.* **1867**(10), 166172 (2021).
- Robinson, M. D. & Oshlack, A. A scaling normalisation method for differential expression analysis of RNA-seq data. *Genome Biol.* **254**, 11 (2010).
- Chen, F. *et al.* RNA-seq analysis identified hormone-related genes associated with prognosis of triple negative breast cancer. *J. Biomed. Res.* **34**(2), 129–138 (2020).
- Fiannaca, A. *et al.* miRTissue (ce): Extending miRTissue web service with the analysis of ceRNA-ceRNA interactions. *BMC Bioinform.* **21**(Suppl 8), 199 (2020).
- Jayarathna, D. K. *et al.* Identifying complex lncRNA/pseudogene-miRNA-mRNA crosstalk in hormone-dependent cancers. *Biology* **10**, 10 (2021).
- Jeggari, A., Marks, D. S. & Larsson, E. miRcode: A map of putative microRNA target sites in the long non-coding transcriptome. *Bioinformatics* **28**(15), 2062–2063 (2012).
- Li, J. H. *et al.* starBase v2.0: Decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* **42**(Database issue), 92–97 (2014).
- Li, R. *et al.* GDCRNATools: An R/bioconductor package for integrative analysis of lncRNA, miRNA and mRNA data in GDC. *Bioinformatics* **34**(14), 2515–2517 (2018).
- List, M. *et al.* Large-scale inference of competing endogenous RNA networks with sparse partial correlation. *Bioinformatics* **35**(14), i596–i604 (2019).
- Yu, G. *et al.* clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **16**(5), 284–287 (2012).
- Therneau, T. M. & Grambsch, P. M. *Modeling Survival Data: Extending the Cox Model* (Springer, 2000).
- Lever, J. *et al.* CancerMine: A literature-mined resource for drivers, oncogenes and tumor suppressors in cancer. *Nat. Methods* **16**(6), 505–507 (2019).
- Ideker, T. *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**(11), 2498–2504 (2003).
- Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**(1), 27–30 (2000).
- Yang, L. *et al.* Low expression of TRAF3IP2-AS1 promotes progression of NONO-TFE3 translocation renal cell carcinoma by stimulating N(6)-methyladenosine of PARP1 mRNA and downregulating PTEN. *J. Hematol. Oncol.* **14**(1), 46 (2021).
- Zhang, X. *et al.* Downregulation of KDR expression induces apoptosis in breast cancer cells. *Cell Mol. Biol. Lett.* **19**(4), 527–541 (2014).
- Jahroudi, N. & Greenberger, J. S. The role of endothelial cells in tumor invasion and metastasis. *J. Neurooncol.* **23**, 99–108 (1995).
- Xiang, T. *et al.* The novel ZEB1-upregulated protein PRTG induced by *Helicobacter pylori* infection promotes gastric carcinogenesis through the cGMP/PKG signaling pathway. *Cell Death Dis.* **12**(2), 150 (2021).
- Chen, C. *et al.* A possible tumor suppressor role of the KLF5 transcription factor in human breast cancer. *Oncogene* **21**(43), 6567–6572 (2002).
- Tong, D. *et al.* Expression of KLF5 is a prognostic factor for disease-free survival and overall survival in patients with breast cancer. *Clin. Cancer Res.* **12**(8), 2442–2448 (2006).
- Qin, S., Ma, F. & Chen, L. Gene regulatory networks by transcription factors and microRNAs in breast cancer. *Bioinformatics* **31**(1), 76–83 (2015).
- Amri, W. A., *et al.* MUC17 and PCNX1 as mediators of chemotherapy response in breast cancer. In San Antonio Breast Cancer Symposium. 2019. Texas.
- Jung, J. W. *et al.* An Activin A/BMP2 chimera, AB215, blocks estrogen signaling via induction of ID proteins in breast cancer cells. *BMC Cancer* **14**, 549 (2014).
- Kui, L. *et al.* High-throughput in vitro gene expression profile to screen of natural herbals for breast cancer treatment. *Front. Oncol.* **11**, 684351 (2021).
- Chen, W. C. *et al.* Systematic analysis of gene expression alterations and clinical outcomes for long-chain acyl-coenzyme a synthetase family in cancer. *PLoS One* **11**(5), e0155660 (2016).
- Jain, P. B. *et al.* The spectraplaklin Dystonin antagonises YAP activity and suppresses tumorigenesis. *Sci. Rep.* **9**(1), 19843 (2019).
- Nair, V. A. *et al.* Oncogenic potential of bisphenol A and common environmental contaminants in human mammary epithelial cells. *Int. J. Mol. Sci.* **21**, 10 (2020).
- Ghodke, I. *et al.* AHNK controls 53BP1-mediated p53 response by restraining 53BP1 oligomerisation and phase separation. *Mol. Cell* **81**(12), 2596–2610.e7 (2021).

39. Mehdi, S. *et al.* LY75 ablation mediates mesenchymal-epithelial transition (MET) in epithelial ovarian cancer (EOC) cells associated with dna methylation alterations and suppression of the Wnt/ β -catenin pathway. *Int. J. Mol. Sci.* **21**, 5 (2020).
40. Park, J. W. *et al.* AHNAK loss in mice promotes type II pneumocyte hyperplasia and lung tumor development. *Mol. Cancer Res.* **16**(8), 1287–1298 (2018).
41. Thomas, M. *et al.* Invasive lobular breast cancer: A review of pathogenesis, diagnosis, management, and future directions of early stage disease. *Semin. Oncol.* **46**(2), 121–132 (2019).
42. Li, N. *et al.* MicroRNA-106b targets FUT6 to promote cell migration, invasion, and proliferation in human breast cancer. *IUBMB Life* **68**(9), 764–775 (2016).
43. Li, J. Y. *et al.* Differential distribution of microRNAs in breast cancer grouped by clinicopathological subtypes. *Asian Pac. J. Cancer Prev.* **14**(5), 3197–3203 (2013).
44. Augoff, K. *et al.* miR-31 is a broad regulator of β 1-integrin expression and function in cancer cells. *Mol. Cancer Res.* **9**(11), 1500–1508 (2011).
45. Lu, Q., Guo, Z. & Qian, H. Role of microRNA-150-5p/SRCIN1 axis in the progression of breast cancer. *Exp. Ther. Med.* **17**(3), 2221–2229 (2019).
46. Parashar, D. *et al.* miRNA551b-3p activates an oncostatin signaling module for the progression of triple-negative breast cancer. *Cell Rep.* **29**(13), 4389–4406.e10 (2019).
47. Huang, Z. M. *et al.* MicroRNA-26a-5p inhibits breast cancer cell growth by suppressing RNF6 expression. *Kaohsiung J. Med. Sci.* **35**(8), 467–473 (2019).
48. Jiang, H. *et al.* MiR-101-3p and Syn-Cal14.1a synergy in suppressing EZH2-induced progression of breast cancer. *Oncol. Targets Ther.* **13**, 9599–9609 (2020).
49. Greenlee, J. D. *et al.* Rafting down the metastatic cascade: The role of lipid rafts in cancer metastasis, cell death, and clinical outcomes. *Cancer Res.* **81**(1), 5–17 (2021).
50. Jayarathna, D. K. *et al.* Integrative transcriptome-wide analyses uncover novel risk-associated microRNAs in hormone-dependent cancers. *Front. Genet.* **20**, 26 (2021).
51. Zhou, S. *et al.* Systematical analysis of lncRNA-mRNA competing endogenous RNA network in breast cancer subtypes. *Breast Cancer Res. Treat.* **169**(2), 267–275 (2018).
52. Jayarathna, D. K. *et al.* A supervised machine learning approach identifies gene-regulating factor-mediated competing endogenous RNA networks in hormone-dependent cancers. *J. Cell. Biochem.* **20**, 20 (2022).

Acknowledgements

The datasets reported in this work originated from the TCGA Research Network: <https://www.cancer.gov/tcga>. The computational resources and services used in this work were provided by the eResearch Office, Queensland University of Technology, Brisbane, Australia.

Author contributions

D.K.J., N.S.G., and M.E.R. designed the study. D.K.J. conducted statistical analysis and data interpretation. D.K.J. prepared the manuscript. N.S.G., M.E.R, and J.B. supervised the study. All authors reviewed the manuscript.

Funding

This research was supported by QUT postgraduate research allowance (QUTPRA); QUT HDR tuition fee sponsorship; Advance Queensland Industry Research Fellowship; the NHMRC Career Development Fellowship; and the Cancer Council Queensland Grant.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-29585-x>.

Correspondence and requests for materials should be addressed to N.S.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023